Information Quality Cluster

Presentations from Breakout Session
"Conveying Information Quality: Recent Progress"

ESIP Summer Meeting
July 16, 2019
Tacoma, WA

Summer 2019 ESIP Meeting – Information Quality Cluster

Topic	Authors	Speaker
Introduction	H. K. Ramapriyan, Ge Peng, David Moroni	H. K. Ramapriyan
NASA Data Quality Working Group's Recommendations and Publications	Yaxing Wei, H. K. "Rama" Ramapriyan, David Moroni, Robert R. Downs, Zhong Liu, Donna J. Scott, and Many WG Members	Yaxing Wei
Uncertainty White Paper Status	David Moroni, H. K. Ramapriyan, Ge Peng	David Moroni
Data Quality Domain Working Group (DQ DWG)@Open Geospatial Consortium (OGC)	Ivana Ivánová	Ivana Ivánová
Update on Maturity Matrix Related Activities	Ge Peng	Ge Peng

NOTE: Author affiliations are provided in the respective presentation slides.

ESIP Information Quality Cluster

Hampapuram “Rama” Ramapriyan[1], Ge Peng[2], David Moroni[3]

[1]Science Systems and Applications, Inc. & NASA Goddard Space Flight Center

[2]North Carolina State University, Cooperative Institute for Satellite Earth System Studies - (CISESS) at NOAA’s National Centers for Environmental Information (NCEI)

[3]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

Breakout Session – Conveying Information Quality: Recent Progress
ESIP Summer Meeting – July 16, 2019
Tacoma, WA

This work was a result of the authors' participation in the ESIP IQC. Ramapriyan's work was supported by NASA under a contract with Science Systems and Applications, Inc. Peng's work was supported by NOAA under a grant with NCSU. Moroni's work was supported by NASA under a contract with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA. U.S. government sponsorship is acknowledged.

Summer 2019 ESIP Meeting – IQC Session Agenda

Topic	Speaker	Minutes
Introduction	H. K. Ramapriyan	10
NASA Data Quality Working Group's Recommendations and Publications	Yaxing Wei	20
Uncertainty White Paper Status	David Moroni	15
OGC and Data Quality (tentative title)	Ivana Ivanova	15
Maturity Matrices Update	Ge Peng	15
Discussion (Identify session's takeaways)	All	15

Meeting notes: http://bit.ly/IQC_20190716_Notes

ESIP Information Quality Cluster

- Vision
  - Become **internationally recognized** as an **authoritative and responsive information resource** for guiding the implementation of **data quality standards and best practices** of the science data systems, datasets, and data/metadata dissemination services.
- Information Quality = {Science Quality, Product Quality, Stewardship Quality, Service Quality}
- What do we do?
  - Share experiences; collaborate internationally; invited speakers at monthly telecons; sessions and/or presentations at AGU, AMS, ESIP, E2SIP, and OGC meetings
  - Maintain wiki site with many useful references http://wiki.esipfed.org/index.php/Information_Quality
- Publications
  - Peng, G. *et al.,* 2016: Scientific stewardship in the Open Data and Big Data era - Roles and responsibilities of stewards and other major product stakeholders. *D.-Lib Magazine,* **22** (5/6), doi: https://doi.org/10.1045/may2016-peng.
  - Ramapriyan, H K, Peng G, Moroni D, Shie C-L, Ensuring and Improving Information Quality for Earth Science Data and Products. *D-Lib Magazine*, **23** (7/8), July/August 2017, DOI: https://doi.org/10.1045/july2017-ramapriyan
  - Moroni, et al.(22 authors), “Understanding and Communicating Uncertainty in Earth Science Data Informatics”, White Paper (in preparation)

Information Quality - Definition

- Scientific quality
  - Accuracy, precision, uncertainty, validity and suitability for use (fitness for purpose) in various applications
- Product quality
  - How well the scientific quality is assessed and documented
  - Completeness of metadata and documentation, provenance and context, etc.
- Stewardship quality
  - How well data are being managed, preserved, and cared for by an archive or repository
- Service Quality
  - How easy it is for users to find, get, understand, trust, and use data
  - Whether archive has people who understand the data available to help users.

IQC Activities

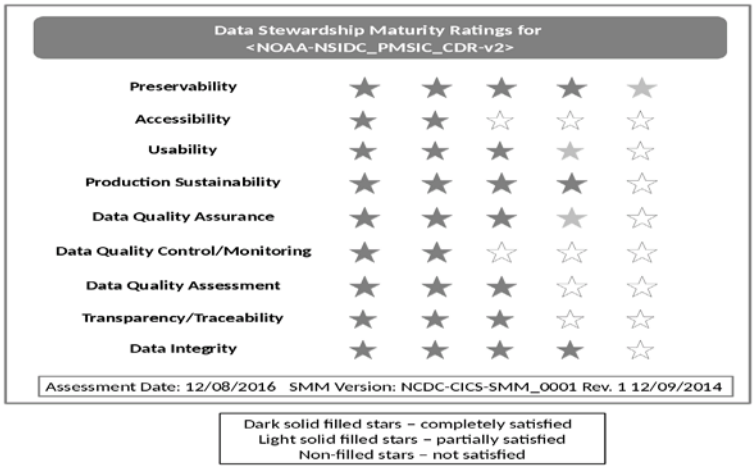
- Share Experiences
  - Collaborate with other ESIP clusters
  - Guest presentations at monthly telecons and biannual meetings
- Actively evaluate best practices and standards for data quality from the Earth science community.
  - Develop and analyze use cases (transitioned from NASA Data Quality Working Group)
- Improve collection, description, discovery, and usability of information about data quality in Earth science data products.
  - Assess recommendations from other groups, e.g., NASA DQWG
- Consistently provide guidance to data managers and stewards on the implementation of data quality best practices and standards as well as for enhancing and improving data maturity.
  - Maturity matrices developed by NOAA and adopted by other organizations
- Support data producers/distributors/intermediaries
  - Provide information about standards and best practices for conveying data quality; mentoring as needed
  - Provide information to help establish, improve, and evolve mechanisms to assist users in discovering, understanding, and applying data quality information properly.
  - Prepare publications
  - Maintain website (wiki) with many useful resources

IQC Activities

Application and Reuse of the NOAA Data Stewardship MM

(Applied to 800+ NOAA Datasets)

- **A part of NCEI *OneStop*-ready process,**
- **Adapted by international data management and stewardship entities** (e.g., CEOS WGISS and WMO CCI)



(Peng et al. Submitted, *Data Science Journal;* Preprint*: bit.ly/DSMM-OneStop*)

White Paper on Earth Science Data Uncertainty

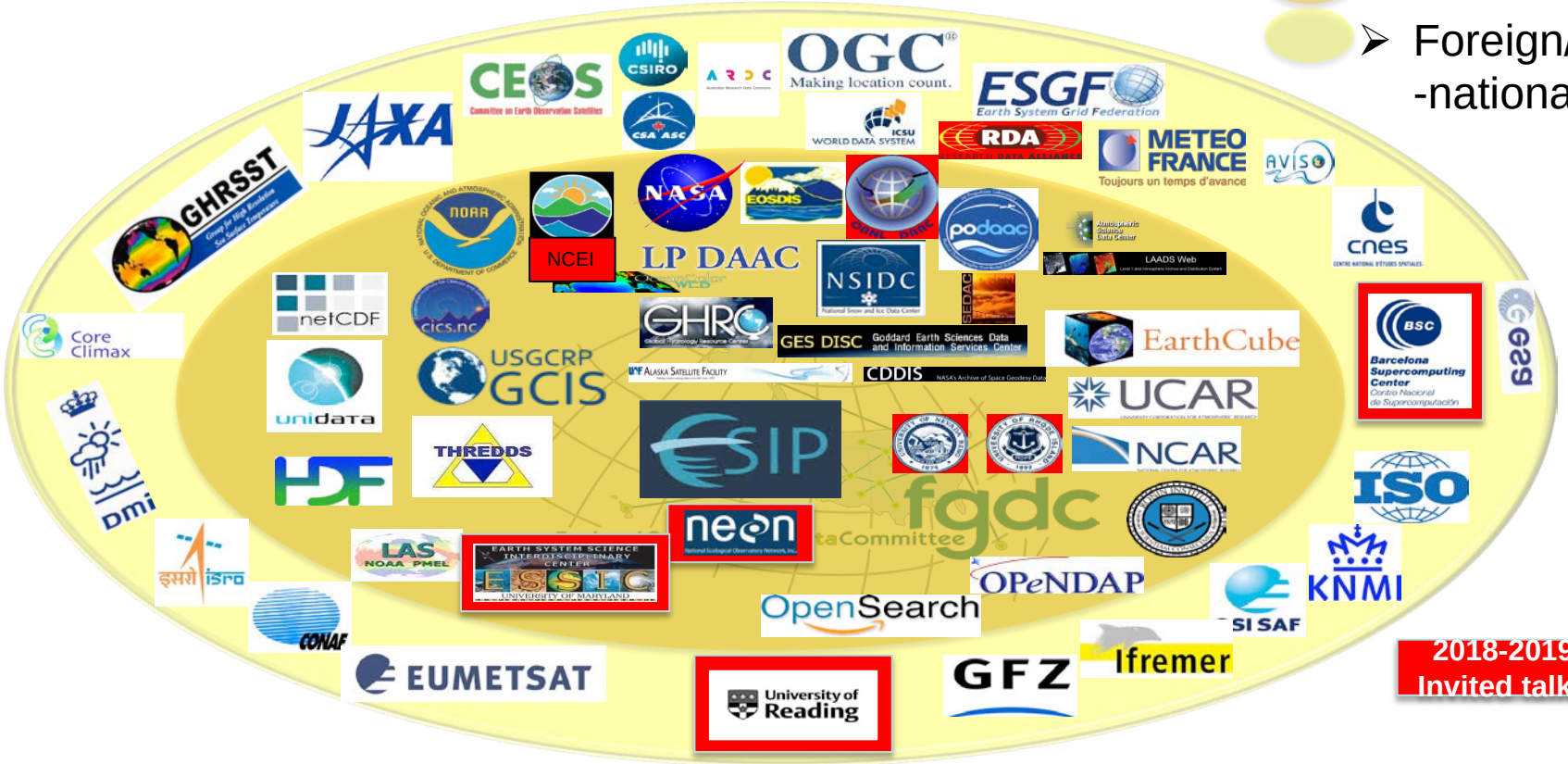
- White Paper in preparation
  - Led by David Moroni
  - ~20 contributing authors
- Focus on “discovery” (not recommendations)
- Document various perspectives about uncertainty
  - Mathematical
  - Programmatic
  - Users’
  - Observational
- Identify commonality and differences between perspectives

IQC Materials – Available to Community

- IQ Wiki Page: **http://wiki.esipfed.org/index.php/**
- Agency Policies and Guidelines
  - **Agency_Policies_on_Information_Quality**
- Relevant Papers: **IQ_Papers**
- IQ Meeting Presentations: **IQ_Presentations**
- Relevant Standards: **IQ_Standards**
- Relevant Web Pages: **IQ_Webpages**
- IQC hosts monthly telecons featuring invited speakers.
  - Presenters have the option of publishing their slides and obtaining a free, citable DOI through the ESIP Figshare.

Many Players Around the World

- US
- Foreign/Inter-national



2018-2019 Invited talks

Summer 2019 ESIP Meeting – IQC Session Agenda

Topic	Speaker	Minutes
Introduction	H. K. Ramapriyan	10
NASA Data Quality Working Group's Recommendations and Publications	Yaxing Wei	20
Uncertainty White Paper Status	David Moroni	15
OGC and Data Quality (tentative title)	Ivana Ivanova	15
Maturity Matrices Update	Ge Peng	15
Discussion (Identify session's takeaways)	All	15

Meeting notes: http://bit.ly/IQC_20190716_Notes

NASA Data Quality Working Group's Recommendations and Publications

Yaxing Wei[1] (*Chair*), H. K. "Rama" Ramapriyan[2] (*Co-chair*), David Moroni[3] (*Former Chair*), Robert R. Downs[4], Zhong Liu[5], Donna J. Scott[6], and *Many WG Members*

[1] ORNL DAAC ; [2] SSAI/GSFC ESDIS; [3] JPL / PO.DAAC; [4] SEDAC; [5] GES DISC; [6] NSIDC;

ESIP Summer Meeting
16 July 2019

Outline

- Data Quality Working Group (DQWG) Outcomes in 2014-2018
  - Data Quality: Aspects, Phases, and Primary Focus Areas
  - Recommendations
  - Implementation Solutions
  - Reuse Readiness Assessment
  - Data Call Template
- 6 Documents Delivered to Earth Science Data and Information System Project (ESDIS) Standards Office (ESO)
  - DMP Template for Data Producers
  - DMP Template for DAACs
  - Comprehensive DQ Recommendations
  - High-priority DQ Recommendations
  - Reuse Readiness Assessment of DQ Software Products
  - Data Call Template and Lessons Learned from the Pilot Study (*in preparation*)

DQWG Outcomes in 2014-2018

Data Quality: Aspects, Phases, and Primary Focus Areas

- Four Aspects
  - Scientific, Product, Stewardship, and Service
- Four Phases of the Data Quality Information Management Lifecycle
  - Capturing; Describing; Facilitating Discovery; and Enabling Use
- Four Primary Focus Areas (of DQWG)
  - Accuracy, Precision and Uncertainty
  - Distinguishability
  - Applicability
  - Usability

Data Quality Recommendations

- 93 Total Recommendations (one example given below)

> 9.4. Develop Tools to Help Users to Leverage Data Quality Information
>
> ESDIS/DAACs should develop tools to help data users easily use data quality information in their research, such as finding, accessing, and processing data based on user-defined quality criteria. For example, all granule level quality metadata should be accessible through clients such as NASA ESDIS' Earthdata Search [5] and Worldview [19], with the highest-level quality description (e.g., good/bad) prominently displayed alongside granule search results or as a layer in visualization tools. Users should also have access to detailed granule level quality information (flags, etc.) as an additional filtering mechanism for subsetting and extraction of quality-specified data.

[5] Earthdata Search, NASA Earthdata web site, 2018: https://search.earthdata.nasa.gov/
[19] Worldview, NASA Earthdata web site, 2018: https://worldview.earthdata.nasa.gov

DQWG, ESDS-RFC-033

Implementation Solutions

- Identified 26 Implementation Solutions in the Master List
  - https://wiki.earthdata.nasa.gov/x/2pASBg

No.	Solution Name	Solution Summary (used to derive relevance)	Implementation Strategy	Benefits of Proposed Implementation Solutions	Solution Point of Contact	Reference URLs
15	NASA GSFC Data Quality Screening Service	A tool developed by @Christopher Lynnes & @Richard Strub for GES-DISC. "DQSS is designed to screen data using both ontology based criteria and user selections of quality criteria (such as minimal acceptable QualityLevel). Data that do not pass the criteria are replaced with fill values, resulting in a file that has the same structure and is usable in the same ways as the original." This service can be utilized before data ingest for the distributor. This service can also be utilized by the public - to further screen the product's quality.	Data quality screening (granule-level filtering)	Provides DAACs a tool to understand quality attributes for overall documentation to product validation. Provides Users a tool to better understand how data decisions regarding quality were established.	@Stacie Doman E	http://opensource.gsfc.nasa.gov/projects/DQSS/

Reuse Readiness Assessment

- Leverage the Reuse Readiness Levels (RRL)

Sol#	Software Name and Version	Software Category	URL	Assessor	Documentation	Extensibility	Intellectual Property Issues	Modularity	Packaging	Portability	Standards Compliance	Support	Verification and Testing	RRL Score
15	Data Quality Screening Service	Data Checker	http://opensource.gsfc.nasa.gov/projects/DQSS/	RRD	3	5	9	7	5	5	3	2	4	4.8
2	Metadata Compliance Checker	Metadata Checker	http://podaac-uat.jpl.nasa.gov/mcc/	RRD	1	1	1	3	1	1	1	6	3	2.0

See Robert R. Downs' Poster

Data Call Template

- Provide a tool for DAACs to evaluate, on a qualitative basis, the overall quality of all varieties of individual Earth science datasets already publicly available. https://wiki.earthdata.nasa.gov/x/BI34Bw

Product Maturity	Cal/Val Documentation Maturity	Quality Flags/Indicators At Each Grid Point	Quality Flags/Indicators Across Multiple Grid Points	Uncertainties are Documented	Biases are Documented	Error Contribution of Input Data
Drop down validation is used to direct input to the proper designation, as defined by this document.	Drop down validation is used to designate the current state that the cal/val documentation is in, whether it is: Preliminary, Technical Report, or Refereed Journal. Preliminary documents include presentations given at science team meetings or white papers that have not been internally vetted by a science team. Technical Reports are published documents that have been vetted by a science team, but have not undergone external peer review. A Refereed Journal covers documents that have gone through external peer review in which the reviewers are selected by the publishing company.	A Boolean (i.e., Yes or No) expression of whether or not these types of quality flags or indicators are provided with the dataset.	Quality Flags/Indicators Exist across Multiple Grid Points (e.g. Global Data File Level or Across a Specified Distribution of Grid Points) - A Boolean (i.e., Yes or No) expression of whether or not these types of quality flags or indicators are provided with the dataset.	A Boolean (i.e., Yes or No) expression of whether or not the uncertainties of dataset have been documented. Insertion of uncertainty information as a data variable and/or the metadata of a dataset constitutes itself as a form of documentation.	A Boolean (i.e., Yes or No) expression of whether or not the biases of dataset have been documented. Insertion of bias information as a data variable and/or the metadata of a dataset constitutes itself as a form of documentation.	This applies only to Level 3 and 4 datasets. Purpose is to provide a high level expression of whether or not the error contribution of the input data is captured somewhere. Selections include "known for all sources", "known for some sources", and "unknown".

Science Quality

Product Quality

...iant	ACDD Compliant	ISO 8601 Compliant	ISO 19115 Compliant	Calibration/Validation Visibility
...d by NASA and ...e recommendations ...ataset Interoperability ...ddition to providing a ...: CF specifications ...ublished an official ...O) document ... for advocating CF across NASA EOSDIS. The primary purpose for DQWG adopting this as a quality-specific attribute is to support usability through interoperability and also to promote standardization with respect to CF representation of quality flags and quality indicators. This field represents a simple Yes/No or N/A Boolean expression to indicate whether compliance is adhered to. N/A is for datasets that are not in netCDF, HDF, or ASCII format. Compliance can be verified by using the PO.DAAC Metadata Compliance Checker (MCC) Tool. The latest CF standard specifications can be found here.	ACDD Conventions are endorsed by NASA and ESDIS as an extension of the recommendations put forth by the ESDSWG Dataset Interoperability Working Group and the ESIP Documentation Cluster. The primary purpose for DQWG adopting this as a quality-specific attribute is to support usability through interoperability and also to promote standardization with respect to ACDD representation of geospatial metadata. This field represents a simple Yes/No or N/A Boolean expression to indicate whether compliance is adhered to. N/A is for datasets that are not in netCDF, HDF, or ASCII format. Compliance can be verified by using the PO.DAAC MCC Tool. More info can be found here.	Represents the date/time standards for metadata in all types of data records and is required for compliance to CF and ISO 19115 (see below). The primary purpose for DQWG adopting this as a quality-specific attribute is to support usability through interoperability and also to promote standardization with respect to ISO representation of date/time metadata This field represents a simple Yes/No or N/A Boolean expression to indicate whether compliance is adhered to. N/A is for datasets that are not in netCDF, HDF, or ASCII format. The PO.DAAC MCC tool automatically checks for this. More info can be found here.	An XML container file and/or Directory Interchange Format (DIF) metadata record (as would be provided to ESDIS for CMR / GMCD / ECHO / REVERB) is extractable and/or provided to ESDIS in this specific geospatial metadata format for the given dataset being evaluated. The primary purpose for DQWG adopting this as a quality-specific attribute is to support usability through interoperability and also to promote standardization with respect to ISO representation of geospatial metadata. This field represents a simple Yes/No Boolean expression to indicate whether compliance is adhered to. More info can be found here.	This is a complementary field to what is being asked for under the Science Quality attribute referred to above as "Cal/Val Documentation Maturity". This field is asking for whether this documentation exists either exclusively at the DAAC, at both the DAAC and some other 3rd party, or whether it is entirely non-existent.

6 Documents Delivered to ESO

[1] DMP Template for Data Producers (*Published*)

- Lead author: *H. K. "Rama" Ramapriyan*
  - *Worked closely with Alfreda Hall (NASA HQ)*
- Provide Data Producers with guidance on the content of DMPs.
- Data Producers - Science Investigator-led Processing System (SIPS), Science Data System (SDS), Research and Analysis (R&A) or Application Program-funded PIs, or a Making Earth System data records for Use in Research for Earth Science (MEaSUREs).
- Includes significant guidance on planning for providing information about data quality.
- https://earthdata.nasa.gov/user-resources/standards-and-references/templates-for-nasa-data-management-plans

[2] Data Management Plan Template for DAACs (*Published*)

- Lead authors: *Donna J. Scott and H. K. "Rama" Ramapriyan*
- Data producers provide DMP covering their products to be delivered to DAACs.
- DAACs may also choose to develop more general DMPs to describe their data operations.
- Template provides DAACs guidance on contents of such DMPs.
- https://earthdata.nasa.gov/user-resources/standards-and-references/templates-for-nasa-data-management-plans

[3] Comprehensive DQ Recommendations for Data Producers and Distributors (*review finished*)

- Lead author: *Yaxing Wei*
- Summarize the approach and outcomes of the DQWG
- Provide a comprehensive set of recommendations regarding data quality that are being offered for producers and distributors of Earth science data.
- 26 existing implementation solutions in the Solution Master List (https://wiki.earthdata.nasa.gov/x/2pASBg)
- Organizations beyond NASA may also benefit from the methodology described here and the resulting recommendations for improvement.

[4] High-Priority DQ Recommendations for Data Producers and Distributors (*Published*)

- Lead author: *Zhong Liu*
- Highlight a subset of recommendations as “actionable and high-priority” for ESDIS to plan and coordinate concrete actions to be taken by data producers and distributors.
- 8 prioritized recommended implementation actions
- Point to existing potential solutions that can be adopted across the EOSDIS and the NASA Earth science research community.
- https://earthdata.nasa.gov/esdis/eso/standards-and-references/recommendations-from-the-data-quality-working-group

[5] Reuse Readiness Assessment of Data Quality Software Products (*In Review*)

- Lead author: *Robert R. Downs*
- Leverage the Reuse Readiness Levels (RRL) developed previously (in 2010) by NASA's Software Reuse WG.
- Provide information to ESDIS and the NASA Earth science research community about the assessment of the reuse readiness of data quality software products that were identified by the DQWG.
- Offer insight into some of the implementation issues that should be considered when planning to adopt software products, as well as a guide for software developers to produce reusable software.

Downs, Ramapriyan, & Wei. A Reusability Assessment of Recommended Software Solutions for Improving the Quality of Earth Science Data Products and Services. IN21B-05 e-Lightning Talk. 2018 AGU Fall Meeting.

[6] Data Call Template and Lessons Learned from the 2017-2018 Pilot Study (*In preparation*)

- Lead author: *David Moroni*
- Data Call Template (https://wiki.earthdata.nasa.gov/x/BI34Bw)
  - Provide a consistent capture mechanism to DAACs for assessing high-level quality characteristics of individual datasets.
- Data Call Pilot Study (https://wiki.earthdata.nasa.gov/x/IYf4Bw)
  - 6 DAACs: ASDC, GES DISC, NSIDC, ORNL DAAC, PO.DAAC, and SEDAC
  - Evaluated the utility of the Data Call Template using DQ info extracted from 14 datasets
- The template is both useful and operationally mature for immediate use.

Many Thanks!!!

Use Cases Contributors

Ed Armstrong (JPL/PO.DAAC), Stacie Doman Bennett (LPDAAC), Lisa Booker (NSIDC), Chris Derksen (U. of Waterloo), Feng Ding (GSFC), Jessica Hausman (JPL), Nathan Kurtz (GSFC), Christopher Lynnes (ESDIS), David Moroni (JPL), H. K. Ramapriyan, Marc Simard (JPL), Vardis Tsontos (JPL)

Writing Authors (Contributors to one or more deliverable documents)

Ed Armstrong (JPL/PO.DAAC), Charlene DiMiceli (UMD), Robert R. Downs (SEDAC/CIESIN), Carolyn Gacke (LP DAAC), Scott Gluck (JPL/CalTech), Ted Habermann, Alfreda Hall (NASA HQ), Beth Huffer (ASDC), George Huffman (GSFC), Siri Jodha Khalsa (NSIDC), Zhong Liu (GES DISC), David Moroni (JPL / PO.DAAC, Former Chair), Byron Peters (SSAI/ESDIS), Hampapuram “Rama” Ramapriyan (SSAI/GSFC ESDIS, Co-Chair), Donna J. Scott (NSIDC), Chung-Lin Shie (GES DISC), Deborah Smith (GHRC), Yaxing Wei (ORNL DAAC, Chair)

Working Group Members

Yaxing Wei (ORNL DAAC, Chair) 2014-**2017-2019**	Larry Di Girolamo (UIUC) 2016-2018	Peter Hall (SSAI/GSFC) 2015-2017	Siri Jodha Khalsa (NSIDC DAAC) 2015-2019	Suhung Shen (GSFC) 2014-2018
David Moroni (JPL/PO.DAAC, Chair) 2014-2017-2019	Feng Ding 2014-2015	Molly Hardman (NSIDC) 2015-2017	Tiffany Matthews (ASDC) 2015-2016	Chung-Lin Shie (GES DISC) 2014-2019
Hampapuram Ramapriyan (SSAI, GSFC/ESDIS, Co-Chair) 2014-2019	Robert R. Downs (SEDAC) 2014 - 2019	Lindsey Harriman (LP.DAAC) 2017-2018	Andrew Mitchell (ESDIS) 2016-2017	Marc Simard (JPL) 2014-2016
Ed Armstrong (JPL/PO.DAAC) 2014-2019	Yonsook Enloe (SSAI/ESO) 2016-2017	Beth Huffer (LaRC) 2015-2019	Sydney Neeley (LP.DAAC) 2017-2018	Deborah Smith (GHRC) 2016-2019
Ross Bagwell (ESDIS) 2014–2018	Carolyn Gacke (LP.DAAC) 2017-2019	George Huffman (GSFC) 2015-2019	Steve Olding (ESDIS) 2014-2018	James Tilton (GSFC) 2017-2018
Stacie Doman Bennett (LPDAAC) 2014 - 2018	Scott Gluck (JPL) 2017-2019	Shannon Leslie (NSIDC DAAC) 2016-2017	Byron Peters (SSAI/ESDIS) 2016-2019	**Gilberto Vicente (GSFC, Co-Chair) 2014-2015**
Michelle Butler (NCSA) 2016-2018	Pierre Guillevic (UMD) 2014-2018	Wenhao Li (JPL) 2017-2018	Bill Rossow (GSFC) 2014-2015	Greg Yetman (SEDAC) 2016-2017
Charlene DiMiceli (UMD) 2014-2019	Ted Habermann (The HDF Group) 2014-2019	Zhong Liu (GSFC/GES DISC) 2017-2019	Donna J. Scott (NSIDC DAAC) 2014-2019	

IQC: Uncertainty White Paper Status

David Moroni[1], Hampapuram “Rama” Ramapriyan[2], Ge Peng[3]

[1]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA

[2]Science Systems and Applications, Inc. & NASA Goddard Space Flight Center

[3]North Carolina State University, Cooperative Institute for Satellite Earth System Studies - (CISESS) at NOAA’s National Centers for Environmental Information (NCEI)

Breakout Session – Conveying Information Quality: Recent Progress
ESIP Summer Meeting – July 16, 2019
Tacoma, WA

This work was a result of the authors' participation in the ESIP IQC. Ramapriyan's work was supported by NASA under a contract with Science Systems and Applications, Inc. Peng's work was supported by NOAA under a grant with NCSU. Moroni's work was supported by NASA under a contract with the Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA. U.S. government sponsorship is acknowledged.

Background and Motivation – Why should we care?

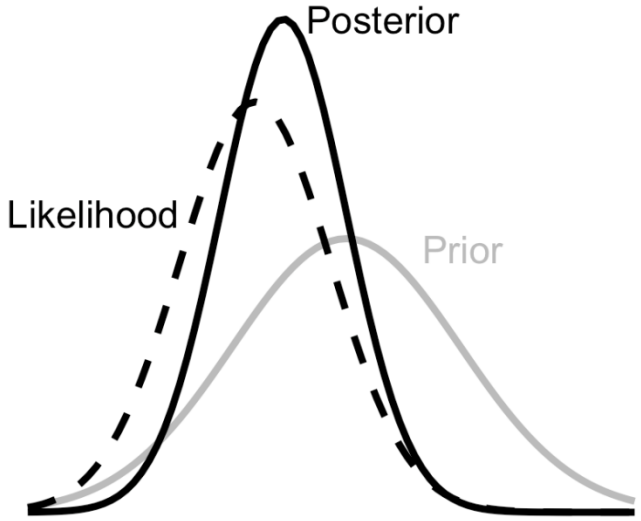
- Uncertainty information provides credibility to data, which leads to credibility of the science, based on such data.
- It can be that unbiased “telephoto lens” into subtleties about data that would otherwise go unnoticed.
- Provides the scientist with discernable information about which dataset is most suitable, in a world where many datasets exist for the same type of observation, based purely on statistics which are agnostic to the results which the dataset(s) may or may not be utilized to support.
- The age of “Big Data” is upon us, but yet many data users (mostly non-experts in numerical analysis) are often left to their own devices as to how to sift through uncertainty information and/or how to derive this information “from scratch”.
- Uncertainty information fundamentally impacts the “science” quality of data, but the availability and packaging of this information can have significant downstream impacts on “product”, “stewardship”, and “service” quality.

White Paper Scope

- Primary focus on “discovery” of the breadth of approaches with regard to Earth science data UQ, UC, and the dissemination/utilization of UQ/UC information by data providers and end users.
- Considers 4 perspectives: Mathematical, Programmatic, Observational, User.
- Will identify both commonalities and differences between perspectives.
- Authors and co-authors represent various aspects of Earth science data informatics, metrology, data science/statistics, remote sensing, in situ, and disciplinary fundamental research.
- Numerical modeling was considered for the sake of use case discussion, but was decided to be left out for the sake of focusing on approaches using observational data.

Mathematical

- Championed by Jonathan Hobbs - JPL
- Considered to be the foundational section of the paper, establishing the key mathematically-based definitions of uncertainty and related constructs such as UQ, UC, mean square error, PDFs, quantiles, confidence intervals, confidence levels, etc…
- Presents directly applicable use cases by which these mathematical definitions are applicable to observational Earth science data, primarily from a remote sensing perspective, but much of which utilizes consistent metrology for a variety of measurement types, including in situ and sub-orbital.



Schematic implementation of Bayes' theorem for a univariate QOI. The prior distribution is combined with information from an observation (via the likelihood) to produce a posterior distribution.

Programmatic

- Championed by Rama – SSAI/NASA GSFC.
- Captures the governmental and intergovernmental approaches, starting with specific US-based agencies and moving into the international arena.
  - Considers US law that drives policy at key agencies, including but not limited to NASA and NOAA.
  - Considers international agreements, such as by the U.N, IPCC, WMO, and CEOS.
  - Considers multi-lateral agreements, statements and policies by EU-sponsored agencies/organizations, such as by: ESA, FIDUCEO, UncertWeb, and MetEOC.

Observational

- Championed by Justin Goldstein – NOAA.
- Discusses the foundational approaches to UQ and UC from an Earth observation perspective, including perspectives from both point-based studies, invariant in space but not in time (e.g., Eulerian Specifications), and those that conduct observations varying in *both* space and time (e.g., Lagrangian Specifications).
- Cal/Val: looks at UQ and UC approaches from a calibration and validation perspective and the role played by “ground truth” data.
- Product Development: examines a variety of approaches and considerations toward making uncertainty information available for common types of observational data products, with a focus on making this information available at the production stage of data.

User

- Championed by Bob Downs – Columbia University.
- Focuses on the ways in which uncertainty information can be effectively or ineffectively consumed, interpreted and ultimately leveraged by the typical data user.
- Provides insights in to methods of communication, dissemination, visualization tools/services, and multi-variate analysis.
- Examples considered include: ISO-19157, UncertML, CO2SYS, and OGC's Testbed-12 innovation program (OGC, 2017).

Next Steps

- Complete by August:
  - Commonalities, differences, conclusions.
  - Re-write the introduction to better align with main sections.
  - Include more graphics/figures.
- Complete by September
  - Prep for white paper publication; consult with Rose Borden to apply improved styling and consistent references/citation styling adhering to AGU standard.

Ideas beyond this publication…

- Draft and publish a shortened “executive summary” paper in a more prominent journal, such as Data Science or EOS.
- Draft a part-2 paper, focusing on recommendations and actionable solutions.

Co-author Acknowledgements (19)

Jonathan Hobbs[1], Robert Wolfe[4], Chung-Lin Shie[5], Christopher J. Merchant[6], Janae Csavina[7], Mark Bourassa[8], *Isla Simpson*[9], Jessica L. Matthews[3], *Matthew Plumlee*[10], Peter Cornillon[11], Justin C. Goldstein[12], Lucy Bastin[13], Kenneth Kehoe[14], Benjamin Smith[15], Jeffrey L. Privette[15], Robert R. Downs[16], Aneesh C. Subramanian[17], Otis Brown[3], Ivana Ivánová[18]

Co-Author Affiliations

1. Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA
2. Science Systems and Applications, Inc. & NASA Goddard Space Flight Center
3. North Carolina State University (NCSU), Cooperative Institute for Climate and Satellites-North Carolina (CICS-NC) at NOAA's National Centers for Environmental Information (NCEI), Asheville, NC
4. NASA Goddard Space Flight Center
5. NASA Goddard Space Flight Center and University of Maryland, Baltimore County
6. University of Reading and National Centre for Earth Observation, Reading, UK
7. National Ecological Observatory Network
8. Center for Ocean-Atmospheric Prediction Studies, Florida State University
9. National Center for Atmospheric Research
10. Northwestern University
11. University of Rhode Island
12. Riverside Technology, Inc. supporting the NOAA NESDIS Technology, Planning, and Integration for Observation division
13. Joint Research Centre, European Commission / Aston University, UK
14. University of Oklahoma- Cooperative Institute for Mesoscale Meteorological Studies
15. NOAA's National Centers for Environmental Information
16. Center for International Earth Science Information Network, Columbia University
17. University of Colorado Boulder
18. Curtin University, Australia

Data Quality Domain Working Group (DQ DWG) @Open Geospatial Consortium (OGC)

Dr Ivana Ivánová
Curtin University
co-chair DQ DWG @ OGC

OGC®
Making location count.

2019 ESIP Summer Meeting, 16 July 2019

Open Geospatial Consortium®

- The Open Geospatial Consortium (OGC) is an international industry consortium of over **532** companies, government agencies and universities participating in a consensus process to develop publicly available interface standards.

- de-facto standardization:
  - “the standards empower technology developers to make complex spatial information and services accessible and useful with all kinds of applications.”
  - OGC standards support interoperable solutions that "geo-enable" the Web

Data Quality Domain Working Group – mission

- WG charter approved in December 2006
- This was in time when ISO DQ standards (ISO 19113, 19114, ISO 19115, ISO 19131 and ISO 19131) existed from 2-6 years.

- WG's **first objective** was to learn what (standardized) DQ means to the various stakeholders in a geospatial supply chain → **DQ Survey in 2007-2008:**
  - 728 respondents (incl. 17% OGC members), both suppliers and consumers of spatial data, from governmental and military organisations, professional, scientific and technology industry (mostly from EU and NA)
- Results:
  - Almost all respondents stated that **data quality is important**, but more than 60% had **no clear approach for managing it** (just some reasons: unaware of standards, missing metadata…).

DQ Survey – impact on the work of DQ DWG in 2008

- Survey resulted in redefining DQ DWG's ('**producer-centric**') mission as:
  - "… a forum for describing an interoperable framework or model for OGC Quality Assurance measures and Web Services to enable access and sharing of high quality geospatial information…"
  - With reference to DQ standards discuss, define and guide in DQ aspects:
    - Accuracy (spatial, thematic and temporal),
    - **Consistency and Integrity**,
    - Completeness,
    - Semantic Interoperability (definitions, languages),
    - Scale, Spatial Reference Systems and Projection.
- WG's participation in OWS4's Geoprocessing Workflow thread – deliverable: Topology Quality Assessment Interoperability Report (https://portal.opengeospatial.org/files/?artifact_id=21821)

DQ DWG @ OGC in 2019

- Currently 41 members from academia, governmental agencies, mapping agencies and private companies

- Co-chairs (as of December 2017):
  - Matt Beare (Beare Essentials)
  - Sam Meek (Helyx secure information systems ltd)
  - Ivana Ivánová (Curtin University)

- Closely related OGC groups:
  - Quality of Service and Experience DWG
  - Citizen Science DWG

DQ DWG program of work

- DQ DWG members contribute to OGC standards:
  - Geospatial User Feedback (http://www.opengeospatial.org/standards/guf)
- DQ DWG members drive discussion on quality:
  - UncertML OGC Discussion paper (https://portal.opengeospatial.org/files/?artifact_id=33234)

- DQ DWG members review standards and engineering reports:
  - ISO 19157, ISO 19115-1 in 2010, ISO 19165-2 in 2019…
  - Testbed 13 in 2017-2018 on reports which implemented 19157 DQ model:
    - FA001: (Aviation) Abstract Quality Model ER
    - FA002: (Aviation) Data Quality Specification ER
    - FA003: (Aviation) Quality Assessment Service ER

- DQ DWG meets at OGC TCs – open forum for discussions on quality

DQ DWG – change of mission

- In December 2017 DQ DWG initiated a revision of the DQ Mission to shift the discussion from ‘producer-centric’ to ‘**user-centric**’.

- Some of the current questions:
  - *What is the understanding of quality within **producers and consumers** of spatial data?*
  - *What are the limitations in **use** of data quality information?*
  - *How to manage the quality of ‘**non-authoritative spatial data**’ to ensure these are **fit for decision-making**?*
  - *How to improve DQ **standards to address fitness for use**?*

NAD ad-hoc – a result of new course of action

- Joint venture of Citizen Science DWG and Data Quality DWG initiated in December 2017 and has met 3 times since

- Currently working on best practice document (https://github.com/opengeospatial/crowdsourcing-vgi/blob/master/ucr.md) for handling crowdsourcing and volunteered geographic information – this will include:
  - advice on metadata that needs to be collected to support the use of crowdsourcing and volunteered geographic information for decision making;
  - advice on how to combine the data and its metadata to facilitate delivery;
  - advice on, or possibly definitions of, RESTful APIs used for receiving crowdsourcing and volunteered geographic information.

DQ DWG stresses collaboration on Spatial Data Quality needs

- Within OGC:
  - Citizen Science DWG (http://www.opengeospatial.org/projects/groups/citizenscience)
  - Quality of Service and Experience DWG (http://www.opengeospatial.org/projects/groups/qosedwg)

- With external partners:
  - ESIP IQC (http://wiki.esipfed.org/index.php/Information_Quality)
  - E2SIP – AU&US collaboration in Earth and Environment Sciences Informatics (no website – summary of latest meeting https://www.esipfed.org/collaboration-updates/2019-e2sip-workshop-summary) – discussion on formation of Australian DQ Interest Group
  - …?

THANK YOU!
ivana.ivanova@curtin.edu.au

OGC®

Update on Maturity Matrix Related Activities

Ge Peng, PhD

North Carolina Institute for Climate Studies (NCICS), NC State University at NOAA's National Centers for Environmental Information (NCEI)

IQC Session: Conveying Information Quality – Recent Progress
ESIP 2019 Summer Meeting, Tacoma, WA, July 16, 2019

Outline

- Application/Adaptation of a Data Stewardship Maturity Matrix (DSMM);
- Status of the RDA FAIR Maturity Matrix;
- Other MM Related Activities

Definitions

Quality – a distinctive attribute or characteristic possessed by something

In This Talk: Data and Information Quality

Maturity – The state of being mature

Maturity Matrix – A maturity assessment model with desired evolution in discrete, progressive stages from a more ad hoc approach to a more managed process

(Based on: **Becker et al.** 2009, *Business & Information Systems Engineering*)

Data Stewardship Maturity Matrix (DSMM)

NCEI/CICS-NC DSMM – Maturity of Stewardship Practices

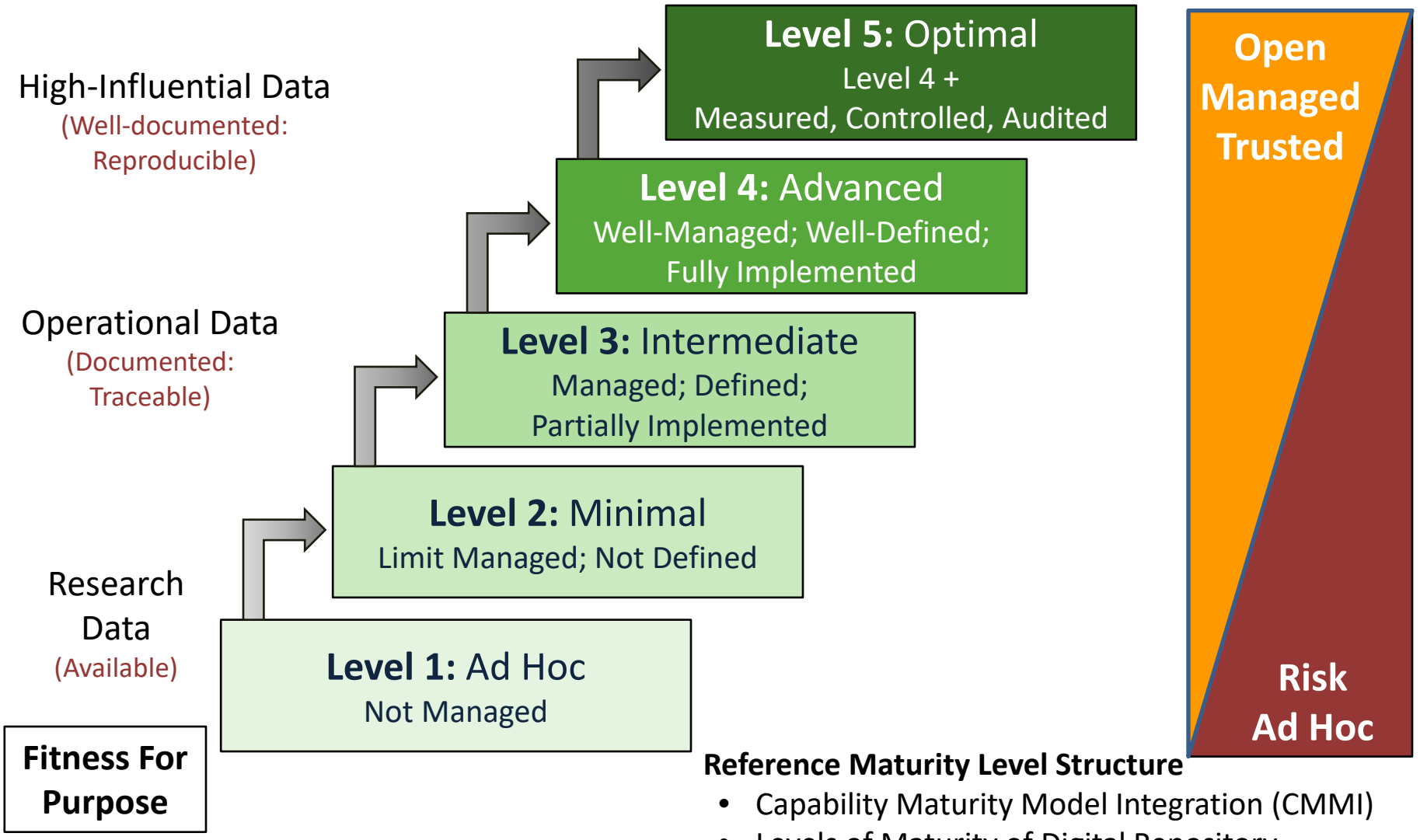
- **Developed Jointly** by domain Subject Matter Experts (e.g., data management, science, and technology),
- **leveraged** institutional knowledge, community best practices, and national and international standards.

Nine Key Components – Open Archival Information System

- Preservability
- Accessibility
- Usability
- Production Sustainability
- Data Quality Assurance
- Data Quality Control/Monitoring
- Data Quality Assessment
- Transparency/Traceability
- Data Integrity

(Self-Evaluation Template: **bit.ly/DSMMtemplate**)

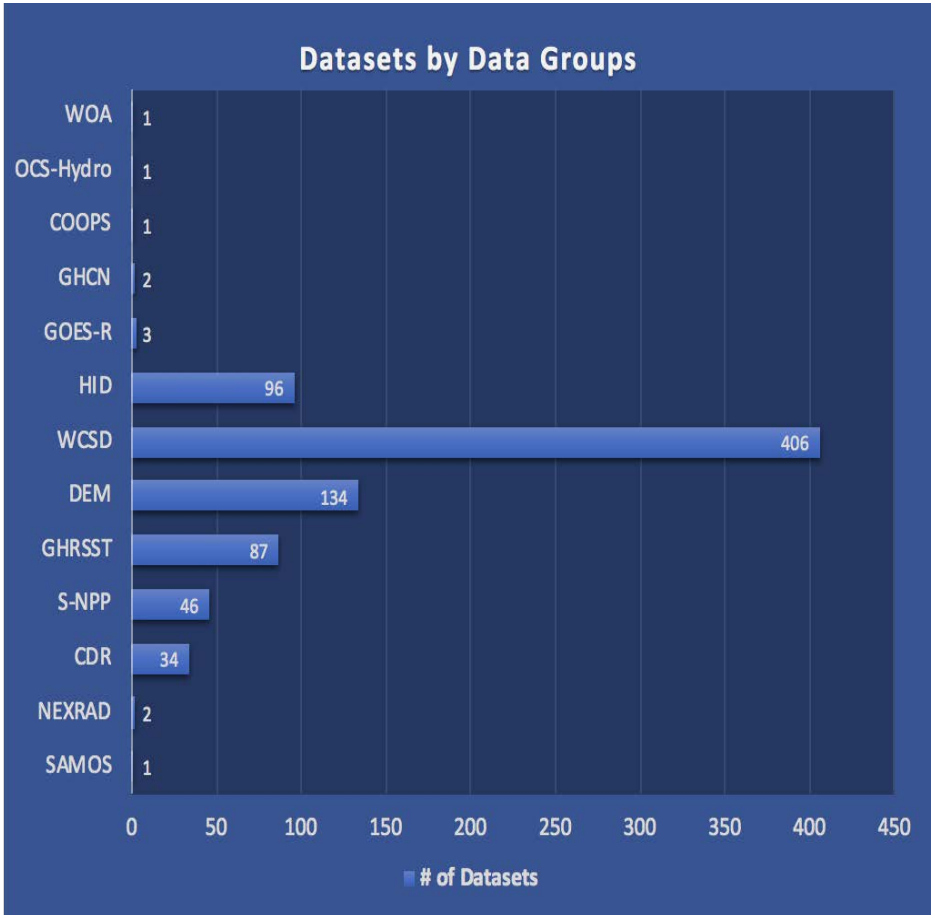
DSMM: Following CMMI Maturity Scale Structure



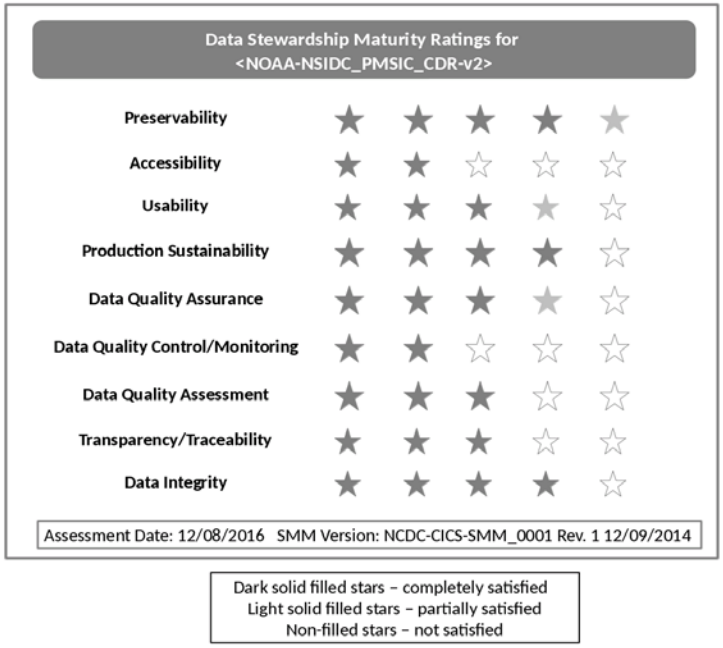
Reference Maturity Level Structure

- Capability Maturity Model Integration (CMMI)
- Levels of Maturity of Digital Repository

NOAA *OneStop* Application of DSMM
(Applied to 800+ NOAA Datasets)



- *OneStop*-ready process,
- NCEI National Archives and Records Administration (NARA) audit,
- NCEI Paleo World Data Center certification.



(Peng et al. 2019, *Data Science Journal, Submitted;* Preprint: *bit.ly/DSMM-OneStop*)

Adaptation of DSMM: CEOS WGISS DMSMM

- The Data Management and Stewardship Maturity Matrix (DMSMM) of the Working Group on Information Systems and Services (WGISS) of the Committee on Earth Observation Satellites (CEOS)
  - Developed by the WGISS Data Stewardship Interest Group, led by ESA
  - Compliant to the implementation Guidelines to the GEOSS Data Management Principles

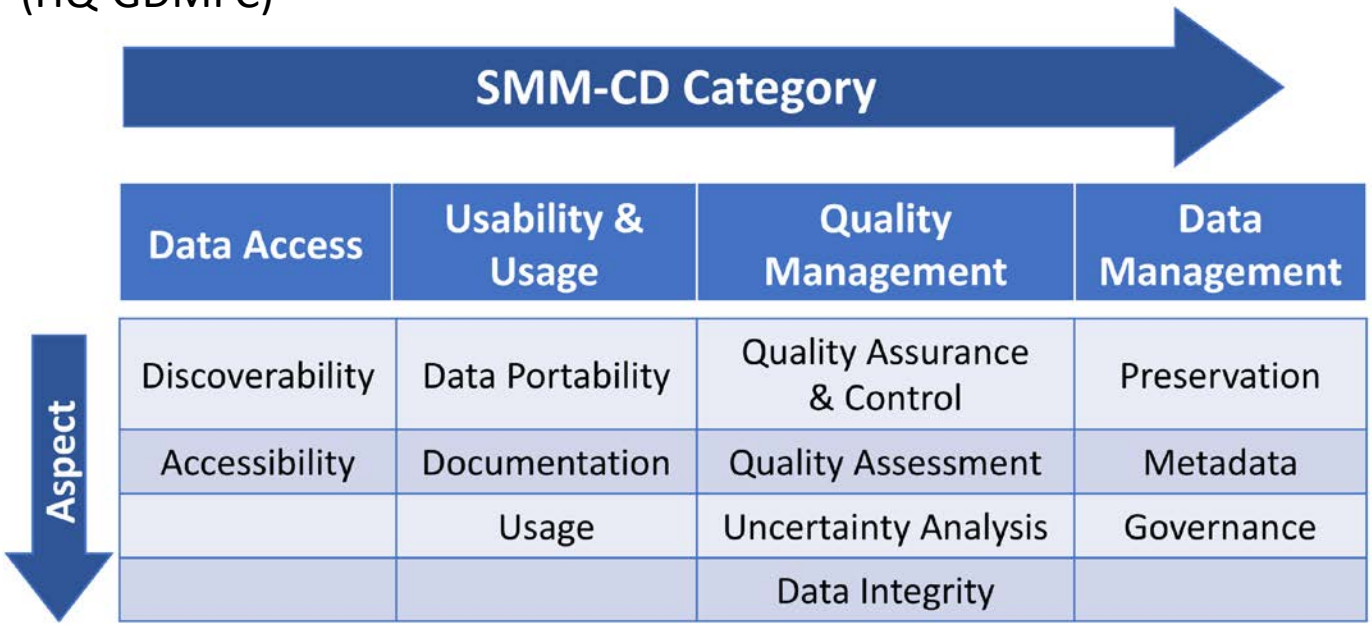
DMSMM	Data Management Principles
Discoverability	DMP-1: Metadata for Discovery
Accessibility	DMP-2: Online Access
Usability	DMP-3: Data Encoding
	DMP-4: Data Documentation
	DMP-5: Traceability
	DMP-6: Quality
Preservation	DMP-7: Preservation
	DMP-8: Verification
Curation	DMP-9: Reprocessing
	DMP-10: Persistent Identifier

More Details - EGU 2019
Poster by Iolando Maggio

https://doi.org/10.6084/m9.figshare.8019350

Adaptation of DSMM: WMO SMM-CD

- The Stewardship Maturity Matrix for Climate Data (SMM-CD) of WMO Commission for Climate
  - Developed by the WMO CCl SMM-CD Working Group,
  - WMO High Quality Global Data Management Framework for Climate (HQ-GDMFC)



Data Access	Usability & Usage	Quality Management	Data Management
Discoverability	Data Portability	Quality Assurance & Control	Preservation
Accessibility	Documentation	Quality Assessment	Metadata
	Usage	Uncertainty Analysis	Governance
		Data Integrity	

(Guidance Booklet: **bit.ly/SMM-CD-Manual**)

Status of FAIR Data Maturity Matrix

FAIR Data Principles:

- Findable,
- Accessible,
- Interoperable,
- Reusable

Wilkinson et al. (2016):

The FAIR Guiding Principles for scientific data management and stewardship

- Started in 2014
- Published on 15 March 2016 by Nature: Scientific Data
- Google: 1432 citations as of Jul 10, 2019

Findable	**F1. (meta)data are assigned a globally unique and persistent identifier**
	F2. data are described with rich metadata (defined by R1 below)
	F3. metadata clearly and explicitly include the identifier of the data it describes
	F4. (meta)data are registered or indexed in a searchable resource
Accessible	**A1. (meta)data are retrievable by their identifier using a standardized communications protocol**
	A1.1 the protocol is open, free, and universally implementable
	A1.2 the protocol allows for an authentication and authorization procedure, where necessary
	A2. metadata are accessible, even when the data are no longer available
Interoperable	**I1. (meta)data use a formal, accessible, shared, and broadly applicable language for knowledge representation**
	I2. (meta)data use vocabularies that follow FAIR principles
	I3. (meta)data include qualified references to other (meta)data
Reusable	**R1. meta(data) are richly described with a plurality of accurate and relevant attributes**
	R1.1. (meta)data are released with a clear and accessible data usage license
	R1.2. (meta)data are associated with detailed provenance
	R1.3. (meta)data meet domain-relevant community standards

RDA FAIR Data Maturity Matrix Working Group

The Goal of the FAIR Data Guiding Principles

- a minimal set of community-agreed guiding principles and practices that allow both machines and humans to find, access, interoperate and re-use research data.

The Implementation Challenges

- No consistent way to implement: interpretation of FAIRness is all over the map,
- Hard to assess FAIRness.

The Goals of the RDA FAIR Data MM WG

- A common set of core assessment criteria,
- Self-assessment model and toolset,
- FAIR data checklist.

(**Source**: https://www.rd-alliance.org/group/fair-data-maturity-model-wg/case-statement/fair-data-maturity-model-wg-case-statement)

RDA FAIR MM WG

- **Timeline:**
  - The Working Group was endorsed by RDA in early January 2019,
  - Delivery in about 12 months (by the EU Commission meeting)

- **Current Status:**
  - Over 100 WG members – strong EU support,
  - Analysis of existing FAIR assessment approaches,
  - Community development of maturity indicators.

RDA FAIR Data MM WG:

https://rd-alliance.org/groups/fair-data-maturity-model-wg

(Source: Case Statement – RDA FAIR Data Maturity Model Working Group; 20190403_FAIR_WG_slides_0.08.pdf)

Other MM Related Activities

- NCEI-ESIP/DSC Use & Services MM
  - Will try again this coming FY (POCs: Peng & Ruth Duerr)
- Synergy of the DSMM and the Core Trustworthy Digital Repository Requirements (CTDRR)
  - NCEI Paleo (POCs: Wendy Gross & Peng)
- MM for model data?!
  - Barcelona Supercomputing Center (the Copernicus Climate Data Store)/ESIP IQC (POCs: Carlo Lacagnina & Peng)

Questions or Comments

Contact Me:

gpeng@ncsu.edu